UNITED STATES PATENT APPLICATION

FOR

METHODS AND APPARATUS FOR
THERMAL MANAGEMENT OF AN INTEGRATED CIRCUIT DIE

INVENTORS:

STEPHEN H. GUNTHER
FRANK BINNS
JACK D. PIPPIN
LINDA J. RANKIN
EDWARD A. BURTON
DOUGLAS M. CARMEAN
JOHN M. BAUER

PREPARED BY:

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN, LLP
12400 WILSHIRE BLVD.
SEVENTH FLOOR
LOS ANGELES, CA 90025-1030

(503) 684-6200

EXPRESS MAIL NO; EV325529755US

METHODS AND APPARATUS FOR
THERMAL MANAGEMENT OF AN INTEGRATED CIRCUIT DIE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001]   This application is a continuation-in-part of U.S. Patent application No. 09/281,237, filed March 30, 1999.

FIELD OF THE INVENTION

[0002]   Thermal management of an integrated circuit die and, more specifically, integrated thermal management systems and methods of use for providing on-chip, closed-loop temperature control at one or more die locations.

BACKGROUND OF THE INVENTION

[0003]   An integrated circuit (IC) device, such as a microprocessor, may include circuitry comprised of many types of discrete circuit components, including transistors, resistors, and capacitors, as well as other components. Semiconductor manufacturers are subject to ever-increasing pressure to increase the speed and performance of such IC devices while, at the same time, reducing package size and maintaining reliability. Thus, by way of example, a modern microprocessor may comprise a die including literally millions of closely-spaced transistors and other discrete components exhibiting sub-micron dimensions and operating at clock rates exceeding 1 GHz. As a result, these microprocessors exhibit high power dissipation and, hence, a corresponding heat load requiring increased cooling capacity, and these high cooling requirements are pushing conventional thermal management technology to its limit.

[0004]   The power consumption of a microprocessor, as well as other types of IC devices, is generally proportional to the operating frequency and the number of transistors required to perform the applications or instructions being executed. The heat generated within a die as a result of this power dissipation must be properly transferred away from the die. If adequate heat transfer does not occur, the die temperature may achieve a level at which performance and reliability suffer or can no longer be guaranteed and, if unchecked, a die temperature may be reached at which permanent structural damage to the microprocessor occurs.

[0005]   The power dissipation of a microprocessor may, however, cause uneven heating of the die. Because one application may primarily utilize one portion of the microprocessor circuitry – i.e., a functional unit, such as an arithmetic and logic unit – while another application may primarily utilize a different functional unit, vastly different amounts of heat may be generated on the die at various locations. Thus, for a particular application, a microprocessor or other IC device may exhibit high temperature locations, or "hot spots," corresponding to die locations proximate one or more functional units experiencing a high workload. Also, with the advent in recent years of microprocessors possessing high speed and performance, there has been increasing disparity between typical power – i.e., the power consumed when running normal applications, such as those run on a personal computer – and maximum power – i.e., the power consumed while running a synthetic, high-power workload specifically designed for maximum power consumption. Thus, although a microprocessor may rarely, if ever, achieve maximum power for an extended period, techniques for thermal management of microprocessors may, in some instances, target the maximum power level to insure that die temperatures will not exceed thermal design limits during operation, potentially causing heat-induced failure or damage to the microprocessor.

[0006]   As is suggested above, thermal management is a critical aspect of the design of modern microprocessors, as well as other IC devices. To remove heat from a semiconductor die, it is known in the art to couple a passive heat transfer device to the die. For example, it is common to thermally couple a heat sink or heat pipe, or other liquid cooling element, to a semiconductor die; however, such passive components possess a limited capacity to dissipate heat. Heat removal may also be facilitated by an active heat transfer device such as a fan, which are often employed in combination with a heat sink having a large surface area (e.g., a plurality of fins). There are, however, several disadvantages associated with the use of fans for cooling IC devices, including poor reliability compared to semiconductor devices, noise, and space requirements. Also, active devices such as fans, as well as the above-described passive device, are generally over-designed for typical power dissipation in order to address the worst case scenario – i.e., the dissipation of maximum power.

[0007]   Another approach to thermal management of an IC die is to actively monitor the temperature of the die. Early thermal monitoring systems consisted of a temperature sensor attached to a heat sink, the heat sink being coupled with a die. If the sensor detected some predetermined threshold temperature, off-chip control hardware and software initiated a response, generally the switching of power to a fan or a reduction in clock frequency. Such thermal monitoring systems are inherently inaccurate due to poor thermal coupling (i.e., a thermal time delay) between the sensor and die. Also, the off-chip control hardware requires the addition of other components to the IC device being sensed or to the next-level assembly, and these added components may consume more surface area in the next-level assembly (i.e., surface area of a circuit board) than the IC device itself.

[0008]   More recently, manufacturers have introduced on-chip thermal sensors that are fabricated directly on a semiconductor die. Although on-chip thermal sensors substantially eliminate the inherent latency of the separately-attached thermal sensor, currently available sensor control and interface logic does not provide reliable temperature measurement and/or closed-loop thermal control. A lack of integration amongst the various elements comprising the conventional sensor control and interface logic, as well as poor integration with the IC device itself, provide a thermal management system exhibiting insufficient response time and, hence, unreliable temperature control.

[0009]   Semiconductor manufacturers have also turned to lower supply voltages to reduce power dissipation of IC devices. However, the increasing speed and circuit density of newer microprocessors will necessitate even lower supply voltages, but the electrical noise generally present in any system inherently limits the degree to which supply voltages may be further reduced.

[0010]   Accordingly, there is a need in the art for an integrated thermal management system and method of use for microprocessors and other IC devices providing reliable on-chip, closed-loop temperature control.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011]   FIG. 1 is a schematic diagram of one embodiment a computer system incorporating a microprocessor having a thermal management system.

[0012]   FIG. 2 is a schematic diagram of one embodiment of the microprocessor shown in FIG. 1.

[0013]   FIG. 3 is a schematic diagram of one embodiment of the thermal management system.

[0014]   FIG. 4 is a schematic diagram of an exemplary embodiment of a temperature sensor for use in the thermal management system shown in FIG. 3.

[0015]   FIG. 5 is an exemplary graph of die temperature as a function of time for a microprocessor having the thermal management system according to FIG. 3.

[0016]   FIG. 6 is a block diagram of one embodiment of a method of performing thermal management.

DETAILED DESCRIPTION OF THE INVENTION

[0017]   The above-noted deficiencies of prior art thermal management techniques are overcome by providing a tightly integrated thermal management system entirely contained on a semiconductor die, such as a microprocessor. Although illustrated below in the context of thermal management of a microprocessor, those of ordinary skill in the art will understand that the methods and apparatus described herein are generally applicable to all types of IC devices.

[0018]   Shown schematically in FIG. 1 is a computer system 5 incorporating a microprocessor 100 providing on-chip, closed-loop thermal management. The microprocessor 100 is coupled via a bus 10 to main memory 20, which may comprise one or more dynamic random access memory (DRAM) devices for storing information and instructions to be executed by microprocessor 100. The main memory 20 may also be used for storing temporary variables or other intermediate information during execution of instructions by microprocessor 100. Computer system 5 also includes read only memory (ROM) 30 coupled via bus 10 to microprocessor 100 for storing static information and instructions for microprocessor 100.

[0019]   The computer system 5 includes one or more input devices 40, such as a keyboard or mouse, coupled to the microprocessor 100 through bus 10. Also coupled to the microprocessor 100 via bus 10 are one or more output devices 50. Typical output

devices 50 include printers and display monitors. The computer system 5 further includes one or more data storage devices 60 coupled to the microprocessor 100 via the bus 10. Common data storage devices include hard disk drives, floppy disk drives, and CD ROM drives. It will be understood by those of ordinary skill in the art that the computer system 5 may include other components and subsystems in addition to those shown and described with respect to FIG. 1. By way of example, the computer system 5 may include video memory, cache memory, as well as other dedicated memory, and additional signal lines and busses.

[0020]   Referring now to FIG. 2, a schematic diagram of microprocessor 100 is shown. Microprocessor 100 comprises a semiconductor die 105 having internal clock circuitry 130 and a plurality of functional units formed thereon, including a bus interface unit 110, an address translation unit 120, instruction decoder unit 140, execution unit 150, and register file unit 160. The microprocessor 100 is powered by a supply voltage (V_{cc}) 102.

[0021]   The schematic diagram of microprocessor 100 presented in FIG. 2 is for illustrative purposes only, and those of ordinary skill in the art will understand that, in practice, a modern microprocessor is generally more complex and may include additional components, such as internal cache, which have been omitted for ease of understanding. Although not shown for clarity, it should also be understand that the internal clock circuitry 130 and various functional units (i.e., bus interface unit 110, address translation unit 120, instruction decode unit 140, execution unit 150, and register file unit 160, as well as internal cache and other systems) of microprocessor 100 are electrically coupled to one another through internal buses and other communication paths. Accordingly, the illustrated microprocessor 100 is presented without limitation, and the apparatus and methods described herein are generally applicable to all types of microprocessors, irrespective of the specific architecture employed.

[0022]   The bus interface unit 110 is coupled to bus 10, as well as main memory 20 and ROM 30. Bus interface unit 110 facilitates transmission of data between main memory 20 and microprocessor 100, and also performs fetching of instructions and other data from ROM 30. Address translation unit 120 performs memory management for microprocessor 100. Specifically, address translation unit stores the memory addresses – whether in main memory 20, internal cache, or other memory – of data being used by the microprocessor 100 during operation. Instruction decoder unit 140 decodes instructions and other control signals received by microprocessor 100.

[0023]   Execution unit 150 is intended to present a broad category of microprocessor functional units comprising a wide range of functions. By way of example, execution

unit 150 may comprise an arithmetic and logic unit for performing arithmetic operations, including shifts, addition, subtraction, multiplication, and division. Register file unit 160 may comprise one or more types of registers for storing data being used by microprocessor 100. For example, register file unit 160 may include integer registers, status registers, instruction pointer registers, and floating point registers, as well as others. If present, internal cache may, for example, be used to store data and control signals from main memory 20.

[0024] External clock 90 provides an external clock signal to internal clock circuitry 130. The internal clock circuitry 130, in turn, provides to microprocessor 100 an internal clock signal derived from the external clock signal. For example, the internal clock circuitry 130 may include a phase lock loop (PLL) circuit that adjusts the frequency of the internal clock signal to a specified multiple of the external clock signal frequency. The internal clock signal that is output from internal clock circuitry 130 is also referred to herein as the system clock for microprocessor 100.

[0025] Microprocessor 100 also includes a thermal management system 200. In one embodiment, thermal management system 200 is formed directly on the die 105 of microprocessor 100 and is entirely contained thereon. Thermal management system 200 is configured to interface with one or more of the functional units on microprocessor 100, as well as internal clock circuitry 130, and to maintain the temperature of die 105 below a specified threshold temperature, the threshold temperature being less than the temperature at which heat-induced damage or operating failures are expected to occur. To measure the die temperature, thermal management system 200 includes at least one temperature sensor 222 formed on die 105 and thermally coupled thereto.

[0026] Shown in FIG. 3 is a schematic diagram of one embodiment of the thermal management system 200. In the illustrated embodiment, the thermal management system 200 comprises four integrated elements: (1) a temperature detection element 220, including at least one thermal sensor 222, as noted above; (2) a power modulation element 240; (3) a control element 260; and (4) a visibility element 280, all of which will be described in greater detail below. It should be understood, however, that other embodiments of a thermal management system 200 may not include all of these elements. For example, a thermal management system 200 may not include a visibility element 280. Further, although the embodiment illustrated herein comprises a thermal management system contained entirely on the die 105 of microprocessor 100, one or more elements or components of the thermal management system 200 may be located off-chip or comprise separate parts attached to the die 105. By way of example, a

thermal management system 200 may include a temperature sensor that is separately fabricated and subsequently attached to an IC device.

[0027]   Shown in FIG. 4 is an exemplary embodiment of a temperature sensor 222 suitable for use with thermal management system 200. The temperature sensor 222 described herein with respect to FIG. 4 is provided for illustrative purposes only, and those of ordinary skill in the art will appreciate that temperature sensor 222 may comprise any other suitable temperature sensor known in the art.

[0028]   The temperature sensor 222 comprises a programmable temperature sensor comprised of circuitry formed directly on the die 105 during manufacture. There exists a number of advantages to forming the sensor 222 directly on die 105. First, improved thermal coupling between the sensor 222 and die 105 is achieved. Second, the temperature sensor 222 can be calibrated at the chip level (as opposed to system level calibration), enabling manufacturing and processing variations from one wafer to the next, and from one section of a wafer to the next, to be nullified.

[0029]   Referring to FIG. 4, temperature sensor 222 includes a programmable voltage 224 and a reference voltage 226, both of which are powered by a current source 228. Sensor 222 further includes a comparator 230, also powered by current source 228. A signal line 237 couples the programmable voltage 224 to comparator 230 and, similarly, a signal line 238 couples the reference voltage 226 to comparator 230, the comparator having an output 239. The reference voltage 226 provides a relatively constant voltage to the comparator 230 over a temperature range of interest. Programmable voltage 224 generates a voltage dependent upon the temperature of die 105 in the vicinity of temperature sensor 222. The comparator 230 compares the programmable voltage 224 against the reference voltage 226 and, if the programmable voltage 224 equals or exceeds the reference voltage 226, the comparator provides a high logic level (i.e., a logical 1) on its output 239; otherwise, the comparator provides a low logic level (i.e., a logical 0) on output 239. The voltage/temperature characteristics of the sensor 222 can be altered via programmable inputs 225 – e.g., a multiplier value – such that the comparator 230 will provide a high logic level at a specified threshold temperature.

[0030]   With reference to FIG. 5, which shows an exemplary curve 300 of die temperature 302 plotted as a function of time 304, the threshold temperature corresponds to a trip point 310 – which will be referred to herein as T_{trip} – selected to insure the temperature of die 105 does not exceed a maximum temperature (i.e., T_{max}) 320 beyond which mechanical and electrical failure of the microprocessor 100 is imminent. To protect against die failure and processor errors, the trip point 310 is generally set at a pre-determined level below T_{max} 320 to insure the maximum thermal condition will not be

achieved. Generally, T_{trip} 310 is set as close as possible to T_{max} 320 to enable microprocessor 100 to operate at or near its highest speed and capacity. In response to detection of T_{trip} 310, the thermal management system 200 will initiate actions, such as engagement of a power reduction mechanism, necessary to lower the die temperature 302 to a second, lower temperature level relative to T_{trip} 310, at which level the comparator 230 of sensor 222 will again assert a low logical level on output 239. This second, relatively lower temperature level, which will be referred to herein as T_{untrip} 330, is set at a temperature differential below T_{trip} 310 to insure all portions of die 105 have sufficiently cooled prior to returning microprocessor 100 to normal operation and, thereby, accounting for any thermal hysteresis within microprocessor 100.

[0031]   In an alternative embodiment, thermal hysteresis may be accounted for by simply engaging, upon detection of T_{trip} 310, a power reduction mechanism for a set time period. After the set time period expires, the temperature sensor 222 is polled to determine if the die temperature still equals or exceeds T_{trip} 310 and, if so, a power reduction mechanism is again engaged for the set time period. The process repeats until, at the end of a time period, the temperature sensor 222 indicates that the die temperature has fallen below the trip point 310, at which time the power reduction mechanism is halted. In a further alternative embodiment, a power reduction mechanism is engaged for an initial time period upon detection of T_{trip} 310. Upon expiration of the initial time period, the temperature sensor 222 is continuously polled for a die temperature measurement. If, at any time thereafter, the temperature sensor 222 indicates that the die temperature has fallen below T_{trip} 310, the power reduction mechanism is disengaged.

[0032]   As noted above, semiconductor die 105 will likely heat unevenly during operation, leading to hot spots on die 105. This uneven heating results from vastly differing application dependent workloads being placed upon the various functional units of microprocessor 100. For example, an arithmetic intensive application may place a high workload on execution unit 150 relative to other functional units, causing high heat generation near execution unit 150. Similarly, a memory intensive application may subject the bus interface unit 110 and/or address translation unit 120 to a relatively high workload, resulting in high temperatures in the vicinity of these functional units, respectively. To compensate for this uneven die heating, the die 105 is thermally mapped to determine the temperature distributions across die 105 during the execution of various types of applications, and the temperature sensor 222 is positioned on the die 105 at a location corresponding to the location of the most severe hot spot – i.e., that location on die 105 achieving the greatest temperature and/or most quickly reaching the greatest temperature.

[0033]   In an alternative embodiment, multiple temperature sensors 222 are provided on microprocessor 100. In this embodiment, each of the multiple temperature sensors 222 is positioned on the die 105 at a location corresponding to a hot spot. If any one of the multiple thermal sensors 222 reaches T_{trip} 310, the thermal management system will take appropriate corrective measures in the vicinity of the tripped sensor or, alternatively, across the entire die 105.

[0034]   Due to errors in sensor placement, as well as sensor inaccuracies, the temperature indicated by temperature sensor 222 may not reflect the true temperature at the sensor location and/or the true temperature of the hottest location on the die 105. For example, if the sensor 222 is erroneously positioned at a location offset from the true location of a hot spot on die 105, the hot spot may reach T_{trip} 310 well prior to the time at which the sensor location achieves this temperature; thus, the output 239 of sensor 222 will lag behind the actual thermal conditions existing at the hot spot. As a result, when the sensor location achieves the T_{trip} 310 level, the temperature at the true location of the hot spot will already have exceeded this temperature level, potentially approaching T_{max} 320. Similarly, sensor errors may result in a failure to timely provide an indication of T_{trip} 310 at the hot spot. To safeguard against heat-induced mechanical or electrical failures, guard bands are used in the selection of T_{trip} 310. A guard band is a specified temperature delta added to (or subtracted from) the trip point to compensate for sensor inaccuracies and location errors. In other words, the trip point 310 is lowered by the temperature delta to insure that, when the sensor 222 indicates T_{trip} 310 (by asserting a high logical level), the true location of a hot spot on die 105 has not actually exceeded the trip point.

[0035]   In another embodiment, as shown in FIG. 4, the temperature sensor 222 may include one or more filters. As noted above, each of the comparator 230 and reference voltage 226 is powered by current source 228. Although current source 228 may be a constant current source, enabling reference voltage 226 to generate a constant voltage level over the temperature range of interest, the power signal from current source 228 may include electrical noise and other transients capable of disrupting operation of temperature sensor 222. For example, small voltage spikes on reference voltage 226 may cause the comparator 230 to signal that the die 105 has achieved the trip point when, in fact, the die has not yet reached this temperature or, likewise, cause the comparator 230 to signal that the die 105 is below the trip point when, in fact, the die 105 is above this temperature. To nullify such voltage spikes, the temperature sensor 222 may include one or more analog filters and/or one or more digital filters.

[0036] Referring to FIG. 4, an analog filter or pulse dampener 232 may be coupled to signal line 238 to dampen out or remove voltage spikes, such that the comparator 230 does not see these transients. In addition to pulse dampener 232 – or, optionally, in lieu of pulse dampener 232 – the temperature sensor 222 may include an analog filter 234 coupled to signal line 237, as well as to signal line 238, and configured to add any voltage spike on reference voltage 226 to programmable voltage 224, such that the respective voltage transients on signal lines 237 and 238 will cancel each other at comparator 230.

[0037] Referring again to FIG. 4, a digital filter 236 may be coupled to the output 239 of comparator 230. Digital filter 236 includes an up-down counter configured to count the number of system clock pulses that occur during a logic state at comparator 230. When the output 239 of comparator 230 transitions from a logical low to a logical high, the up-down counter will count upwards, incrementing once for every clock pulse occurring while the comparator output 239 remains high. When the output 239 of comparator 230 transitions from a logical high back to logical low, the up-down counter will count downwards, decrementing once for every clock pulse occurring while the comparator output 239 remains low.

[0038] As noted above, electrical noise present in reference voltage 226 may cause the comparator 230 to change states when, in fact, an actual corresponding change in temperature conditions has not occurred. The digital filter 236 is configured to provide an output – i.e., a change of state – only if the up-down counter counts a specified number of clock cycles that indicates a "real" temperature event has occurred. If the comparator output 239 is continually transitioning between logical high and logical low due to electrical noise, the up-down counter of digital filter 236 will not increment past the specified number of clock cycles, as the transitions from high to low and from low to high due to noise will essentially cancel out. Thus, by allowing state transitions at the comparator output 239 due to electrical noise to cancel out, the digital filter 236 substantially eliminates the effects of electrical noise in temperature sensor 222. Also, the digital filter 236 can be tuned for various operating conditions and environments by altering the specified number of clock cycles corresponding to a real temperature event.

[0039] Power modulation element 240 (see FIG. 3) comprises control and logic circuitry configured to perform any one or a combination of techniques directed at reducing the power dissipation of microprocessor 100 to lower the temperature of die 105, or to lower the temperature of a specific region thereof. Control and logic circuitry for power modulation element 240 is entirely contained on semiconductor die 105, as noted above.

[0040] The integration of the temperature detection element 220 and power modulation element 240 directly on die 105 provides several advantages. First, the latency between detection of the trip point and the corresponding response of the power modulation element 240 will be low. Second, by substantially reducing latency effects, efficient power reduction is possible. Specifically, a loss in performance may accompany at least a substantially comparable reduction in power consumption or, if a nonlinear relationship between performance loss and power consumption is achieved, a larger reduction in power consumption. For example, if the power modulation element 240 causes a 10% reduction in performance while operating, this element also provides at least a 10% reduction in power consumption, such an ideal relationship generally being possible if clock frequency is the only variable. Third, as the circuitry for power modulation element 240, as well as for temperature detection element 220, is constructed during the normal course of manufacturing semiconductor die 105, the power modulation element 240 (and thermal management system 200 generally) will add only minimal design and manufacturing costs.

[0041] The power modulation element 240 may be configured to implement any one or more of a number of power reduction mechanisms. Such power reduction mechanisms may be broadly classified into those techniques that directly lower the power consumption of microprocessor 100, those techniques that limit the speed at which microprocessor 100 executes instruction, and those techniques that limit the number of instructions executed by microprocessor 100. In addition to these power reduction mechanisms, those of ordinary skill in the art will appreciate that any other suitable power modulation technique may be employed in power modulation element 240.

[0042] The power consumption of microprocessor 100 may be directly reduced by lowering the supply voltage (V_{cc}) 102 of microprocessor 100. Although there are limits to the extent that supply voltage 102 can be reduced – due to electrical noise and other transients, as noted above – any reduction in supply voltage 102 will result in a corresponding decrease in power consumption (i.e., the product of voltage and current).

[0043] There are a number of approaches to reducing the speed at which microprocessor 100 executes instructions and, hence, providing a corresponding reduction in power consumption. As noted above, the power consumption of the microprocessor 100 is proportional to the system clock frequency and the number of transistors required to perform any running applications. One approach comprises simply lowering the system clock frequency provided by internal clock circuitry 130. Diminishing the system clock frequency slows the rate at which microprocessor 100 executes instructions, thereby reducing power consumption.

[0044]   The system clock signal may also be halted – either periodically or for a set period of time – to microprocessor 100 or to one or more specific functional units thereof. The shutting off of the system clock to specific functional units or the entire microprocessor 100 for a set time period is often referred to as clock gating. Periodically interrupting the system clock to one or more functional units at a specified duty cycle is commonly referred to as clock throttling. It should be noted that the terms clock gating and clock throttling are often used interchangeably, and clock gating can be viewed as a special case of clock throttling in which there is only one cycle. By way of example, for an arithmetic intensive application, clock throttling may be applied to the execution unit 150 at a 50% duty cycle (i.e., periodically halting the system clock to the execution unit 150 for a period of time followed by providing the system clock signal thereto for an equal time period).

[0045]   Clock throttling may be achieved my modulating the STOPCLOCK signal to the microprocessor 100. The STOPCLOCK request causes the system clock signal to be stopped to the bulk of a microprocessor's logic for a short time period. The STOPCLOCK request is a relatively high priority interrupt, and the delay between the STOPCLOCK signal and the resulting power decrease is relatively short (e.g., less than 1 microsecond). Initiating clock gating or clock throttling is not limited to use of the STOPCLOCK request, and it should be understood that other methods of performing clock gating or clock throttling may be employed.

[0046]   Alternatively, power consumption of microprocessor 100 may be reduced by selectively blocking clock pulses generated by internal clock circuitry 130. For example, every third clock pulse of the system clock signal may be eliminated to reduce the speed at which microprocessor 100 executes instructions. The partially blocked clock signal may be provided to only a portion of the microprocessor 100 or to the entire microprocessor 100.

[0047]   Limiting the number of instructions executed by the microprocessor 100 may be achieved by selectively disabling one or more functional units thereof or, alternatively, by selectively limiting instructions sent to one or more functional units. As an example of the first approach, during an arithmetic intensive application, operation of the execution unit 150 may be temporarily halted to cool die 105 in the vicinity of execution unit 150. As an example of the latter approach, again during an arithmetic intensive application, only selected instructions are delivered to execution unit 150, while delivery of other instructions is temporarily halted, to allow die 105 to cool. In a further alternative embodiment, rather than selectively disabling one or more functional units of

microprocessor 100, the power consumption of microprocessor 100 may be reduced by changing the behavior of one or more of its functional units.

[0048]   In an alternative embodiment, the temperature detection element 220 is configured to provide a signal at output 239 of comparator 230 for each of a plurality of trip levels, the highest trip level corresponding to T_{trip} and the other trip levels corresponding to subsequently lower temperatures. Providing multiple trip levels allows power modulation element 240 to perform adaptable power reduction. For example, the plurality of trip levels may correspond to an equal number of programmed power reduction schemes. At the lowest detected temperature, power modulation element 240 would implement one type of power reduction mechanism, such as engaging clock gating or throttling. If the next highest temperature is detected by temperature sensor 222, power modulation element 240 would initiate another power reduction mechanism – for instance, lowering the supply voltage 102 to microprocessor 100. Higher trip levels would result in the performance of other power reduction mechanisms, or combinations thereof. Finally, at the trip point, power modulation element 240 would execute yet another power reduction mechanism, such as the selective disabling of functional units.

[0049]   Generally, the duty cycle of power modulation element 240 is static, and the same duty cycle is used whenever clock throttling is performed. Providing multiple trip levels, however, also allows for use of a programmable duty cycle while performing clock throttling. At the lowest detected temperature, clock throttling may be performed at a first duty cycle and, at the next highest detected temperature, a second duty cycle may be imposed during clock throttling. Power modulation element 240 is similarly programmed with duty cycles for each successively higher trip level, including the trip point. Thus, the duty cycle used during clock throttling may vary dynamically with changes in thermal conditions.

[0050]   Control element 260 (see FIG. 3) comprises one or more registers 265 enabling control over the behavior of thermal management system 200. For example, a register 265 may be configured to provide an enable/disable bit, allowing the thermal management system 200 to be selectively enabled. Other registers 265 may provide the ability to disengage or by-pass specific elements or components of thermal management system 200. By way of example, it may be desirable to disable one or more of the pulse dampener 232, analog filter 234, and digital filter 236 for testing. Also, a register 265 may provide for activation of thermal management system 200 upon the occurrence of some external event. For example, in a multi-processor computer system, when an elevated temperature is sensed in one microprocessor, it may be desirable to engage the thermal management systems of other microprocessors in the system, irrespective of

whether the thermal management system of each of those other microprocessors was enabled. Alternatively, a register 265 may provide a force active bit, enabling the thermal management system 200 to be forced into operation, thereby overriding any enable/disable bit that has been set to disable or any external software that has shut down the thermal management system 200.

[0051] A further aspect of control element 260 is compatibility with the Advanced Configuration and Power Interface (ACPI) specification, enabling control of thermal management system 200 by external software and devices. See *Advanced Configuration and Power Interface Specification*, Revision 2.0, July 27, 2000. The ACPI specification defines a hardware and software environment that allows operating system software complete visibility and control of system configuration and power management. Control element 260 includes at least a register 267 – defined in the ACPI as the "Processor Control" register, or "P_CNT" – that enables clock throttling and can force thermal control to begin. The P_CNT register also provides for the setting of a duty cycle. The incorporation of all elements of thermal management system 200 entirely on microprocessor 100 in combination with the P_CNT register defined by the ACPI specification enables the operating system software running on computer system 5 to initiate thermal management on microprocessor 100. This holds true even when microprocessor 100 comprises one among many microprocessors in a multi-processor system and the system's chipset does not support per processor clock throttling, because thermal management system 200 is contained entirely on microprocessor 100.

[0052] Visibility element 280 (see FIG. 3) comprises a plurality of devices enabling the outside world – e.g., computer system 5 – to monitor the status of thermal management system 200. The output 239 at comparator 230 of temperature sensor 222 provides a signal (either logical high or logical low) than can be monitored by external devices, enabling some action to be taken at the system level if the die temperature reaches the trip point. Similarly, the visibility element 280 may include a register 281 providing a status bit indicating the state of comparator 230, essentially providing the same information as the output 239 at comparator 230. In addition, the visibility element 280 may include a register 282 providing a "sticky" bit. The sticky bit is set the first time the temperature sensor 222 detects a trip point, and the sticky bit must be explicitly cleared by a software command or by a microprocessor reset. Such a sticky bit may, for example, be used by diagnostic software to determine if the die 105 had ever reached the trip point temperature.

[0053] Another visibility mechanism that may be provided by visibility element 280 is the ability to generate an interrupt whenever there is a state change at output 239 of

comparator 230. Such an interrupt can be generated when the thermal sensor 222 transitions to T_{trip} (i.e., from a "not hot" state to a "hot" state) and/or when the thermal sensor transitions to T_{untrip} (i.e., from the "hot" state to the "not hot" state). Each of these interrupts can be individually enabled or disabled by registers 283, 284, respectively.

[0054]   In addition, visibility element 280 may include one or more counters 286 configured to count the number of clock cycles lost as a result of thermal management system 200. The information provided by such a counter 286 may provide a useful tool for a number of purposes. For example, knowledge of lost clock cycles can be used by microprocessor designers for design verification and management. Knowledge of lost clock cycles may be used for thermal management itself. By way of example, if a pre-determined number of clock cycles are counted, the thermal management system 200 may be configured to take additional action to even further reduce power consumption of microprocessor 100. Similarly, counter 286 may be used to generate interrupts when a specified number of lost clock cycles is exceeded. Information from counter 286 may also be used to create an error log for use in system maintenance.

[0055]   Those of ordinary skill in the art will understand that the various functional units of microprocessor 100 – i.e., bus interface unit 110, address translation unit 120, instruction decoder unit 140, execution unit 150, and register file unit 160 – as well as the elements of thermal management system 200 – i.e., temperature detection element 220, power modulation element 240, control element 260, and visibility element 280 – are not necessarily discrete circuit components but, rather, one or more of these functional units or elements may be integrated or share common circuitry. For example, thermal management system 200 and one or more functional units of microprocessor 100 may have common circuitry.

[0056]   The above-described temperature detection, power modulation, control, and visibility elements 220, 240, 260, 280 are integrated into a single thermal management system 200 fabricated directly and contained entirely on die 105 and providing closed-loop thermal control on microprocessor 100. The tight integration amongst these elements and with microprocessor 100 provides a number of advantages. Guard bands are minimized and latency reduced, resulting in improved accuracy and chip-level reliability. By providing improved reliability at the chip level, greater device protection for microprocessor 100 is achieved, thereby providing higher system-level reliability and fewer failures in higher level systems. In addition, greater power control is provided by thermal management system 200; however, there is no comparable increase in manufacturing costs associated with this improved performance and control, as the on-

chip fabrication of thermal management system will not significantly increase manufacturing costs, as noted above.

[0057]   Shown in FIG. 6 is a schematic diagram of one embodiment of a method of thermal management 400. The method of thermal management 400, which may be used for thermal management of, for example, microprocessor 100, may be initiated in control element 260. Denoted at 410, if the enable/disable bit in a register 265 of control element 260 is set to disable, thermal management system 200 is disabled, which is denoted at 415. If the enable/disable bit in the register 265 is set to enable, the thermal management system 200 will switch to active and temperature detection element 220 will commence measurement 420 of the temperature of die 105 at the location of temperature sensor 222 or, optionally, at multiple die locations if more than one sensor 222 is employed. In an alternative embodiment, temperature detection element 220 is always measuring the temperature at one or more die locations, and setting the enable/disable bit to enable simply allows the thermal management system 200 to act on the output 239 of temperature sensor 222 (or the output of digital filter 236).

[0058]   In alternative embodiments, other registers 265 of control element 260 may be employed during thermal management method 400. Registers 265 of control element 260 may be used to disengage or by-pass specific elements or components of thermal management system 200, to cause activation of thermal management system 200 upon the occurrence of some external event, or to force activation of thermal management system 200 in order to override the enable/disable bit or external software (e.g., a command received on the P_CNT register 267). In a further alternative embodiment, external software, such as the operating system running on computer system 5, may activate thermal management system 200 via P_CNT register 267.

[0059]   If the temperature sensor 222 detects T_{trip}, which is denoted at 430 in FIG. 6, power modulation element 240 will be engaged to initiate power reduction 450. Power modulation element 240 may employ any one of a number of power reduction mechanisms to reduce power consumption of microprocessor 100 and lower the temperature of die 105. Power consumption of microprocessor 100 may be reduced by lowering the supply voltage 102 thereto, by lowering the system clock frequency of microprocessor 100, by initiating clock gating or clock throttling of the system clock, by selectively blocking clock pulses generated by internal clock circuitry 130, by disabling one or more selected functional units of microprocessor 100, by selectively limiting instructions sent to one or more functional units of microprocessor 100, by changing the behavior of one or more functional units of microprocessor 100, or by engaging any combination of these mechanisms. In an alternative embodiment of the method of

thermal management 400, in which the temperature sensor 222 is configured to provide an output at comparator 230 for multiple trip points, thermal management system 200 may utilize any one or suitable combination of the above-noted power reduction mechanisms at each of the multiple trip points. In a further alternative embodiment, a programmable duty cycle is employed to provide a specified duty cycle at each of a plurality of trip points.

[0060]   In another embodiment of thermal management method 400, visibility element 280 provides visibility 440 into thermal management system 200. Thermal management visibility 440 may be provided by the output 239 of temperature sensor comparator 230, by a status bit in register 281 indicating the state of comparator 230, by providing a sticky bit in register 282 to indicate the first occurrence of T_{trip}, by the generation of an interrupt when the state of comparator 239 changes, or by a counter 286 indicating the number of clock cycles lost due to thermal management.

[0061]   If T_{untrip} is detected by thermal sensor 222, as denoted at 460 in FIG. 6, any power reduction mechanisms being implemented by thermal management system 200 are halted 470. So long as the enable/disable bit of control element 260 is set to enable, thermal management system 200 continues to monitor the temperature of die 105 at location of temperature sensor 222 and to execute appropriate power reduction measures as required.

[0062]   In an alternative embodiment, power reduction 450 is conducted for a set period of time and, upon expiration of the set time period, the temperature sensor 222 is polled. If the temperature sensor 222 indicates the die temperature is below T_{trip}, power reduction is halted 470. If, however, the temperature sensor indicates that the die temperature is at or above T_{trip}, power reduction 450 is continued for another of the set time periods, and the process repeats until the die temperature falls below T_{trip}. In a further alternative embodiment, after expiration of the first initial time period, the temperature sensor 222 is polled continuously until the die temperature falls below T_{trip}, at which time any power reduction mechanisms are halted 470.

[0063]   The method of performing thermal management 400 of an IC device provides on-chip, closed-loop temperature control at one or more die locations. The method 400 utilizes the integrated elements – temperature detection element 220, power modulation element 240, control element 260, and visibility element 280 – of the thermal management system 200. However, in alternative embodiments, the method of thermal management 400 may proceed without using all of these elements. By way of example, a method of thermal management 400 may not include the provision of visibility 440 (see FIG. 6).

[0064]   The foregoing detailed description and accompanying drawings are only illustrative and not restrictive. They have been provided primarily for a clear and comprehensive understanding of the illustrated embodiments and no unnecessary limitations are to be understood therefrom. Numerous additions, deletions, and modifications to the embodiments described herein, as well as alternative arrangements, may be devised by those skilled in the art without departing from the spirit of the present invention and the scope of the appended claims.